

TÁI NHẬN DẠNG NGƯỜI TRONG HỆ THỐNG NHIỀU CAMERA GIÁM SÁT KHÔNG CHỖNG LẤP

Trần Thanh Toàn^{1,2}, Ngô Đình Phong¹, Hồ Phước Tiến³, Chế Việt Nhật Anh¹ và Trương Công Dung Nghi¹

¹ Khoa Điện-Điện Tử, Trường Đại học Bách Khoa, Đại học Quốc Gia Tp. HCM

² Trường Cao đẳng nghề Đồng Tháp

³ Khoa Điện tử-Viễn thông, Trường Đại học Bách Khoa, Đại học Đà Nẵng

Thông tin chung:

Ngày nhận: 19/09/2015

Ngày chấp nhận: 10/10/2015

Title:

Person re-identification through non-overlapping camera system

Từ khóa:

Tái nhận dạng người, hệ thống camera không chồng lấp, chỉ số tương đồng, chỉ số nổi bật

Keywords:

Person re-identification, non-overlapping camera system, similarity score, saliency score

ABSTRACT

Person re-identification through non-overlapping camera system recently becomes one of the most interesting problems in automated video surveillance. Once people are detected in the scene, they are characterized by the appearance-based models. However, the background in the cropped image obtained from the human detector could still affect the model used for person re-identification. Thus, in this paper, we propose a re-identification system including two main tasks: removing the background in the image of the detected people; and reinforcing the local patches that are discriminative to different people, and reliable in retrieving the same person in different non-overlapping cameras.

TÓM TẮT

Bài toán tái nhận dạng người hướng tới việc tự động theo vết và tìm kiếm lại đối tượng di chuyển trong hệ thống camera giám sát với góc nhìn của các camera không chồng lấp nhau. Khi một đối tượng được phát hiện trong khung ảnh, các đặc trưng về hình dáng bên ngoài của đối tượng sẽ được trích xuất thông qua việc xây dựng các mô hình đặc tả. Tuy nhiên, phần nền còn lại của khung ảnh trích xuất đối tượng thu được từ giải thuật phát hiện người vẫn có thể ảnh hưởng đến mô hình tái nhận dạng người. Vì vậy, trong bài báo này, chúng tôi đề xuất chuỗi xử lý bài toán tái nhận dạng người với hai nhiệm vụ chính: loại bỏ vùng nền trong khung ảnh đối tượng đã được phát hiện và củng cố đặc trưng tách từ các mảnh ghép cục bộ dùng để phân biệt hai đối tượng khác nhau. Việc kết hợp này nhằm nâng cao mức độ tin cậy của việc truy vấn cùng một đối tượng ở hai camera khác nhau với góc nhìn không chồng lấp.

1 GIỚI THIỆU

Tái nhận dạng người trong hệ thống nhiều camera giám sát không chồng lấp ngày càng có nhiều ứng dụng, đặc biệt là trong các hệ thống camera giám sát công cộng. Các camera trong hệ thống này giám sát hàng trăm địa điểm, hàng trăm đến hàng nghìn người mỗi ngày nên việc xác định và tìm kiếm đối tượng cần truy vấn một cách thủ

công là hết sức khó khăn và tốn rất nhiều thời gian, công sức. Do đó, bài toán tái nhận dạng các đối tượng cần truy vấn nhằm hướng tới một hệ thống hoàn toàn tự động ngày càng được chú trọng. Việc tái nhận dạng các đối tượng thường phải đối mặt với nhiều khó khăn như sự thay đổi về ánh sáng, góc quay của camera, sự tương đồng về hình dáng bên ngoài của các đối tượng...

Để thực hiện việc tái nhận dạng người, hầu hết các nghiên cứu trước đây dựa trên các thông tin toàn cục như thông tin về phân bố màu sắc, hướng tính như (B. Prosser *et al.*, 2010; Hirzer Martin *et al.*, 2011; Mert Dikmen, 2011; W. Zheng, 2011; Martin Hirzer *et al* 2012; A. Mignon và F. Jurie, 2012;). Các phương pháp nhận dạng này thường đạt hiệu quả cao khi góc nhìn không có sự thay đổi lớn và đặc điểm hình dáng bên ngoài của các đối tượng là tương đối khác nhau. Tuy nhiên, khi góc nhìn thay đổi đủ lớn, các đặc tính hình dáng có nhiều điểm tương đồng hoặc miền ảnh nền khá giống nhau thì hiệu suất nhận dạng giảm đi đáng kể.

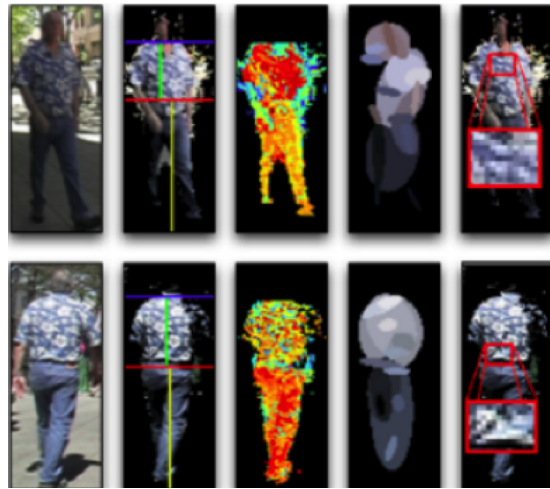
Trong bài báo này, chúng tôi đề xuất chuỗi xử lý bài toán tái nhận dạng người trong hệ thống camera với góc nhìn không chồng lấp gồm hai bước chính. Đầu tiên, với các khung ảnh trích xuất đối tượng thu được từ khối phát hiện người (human detector), chúng tôi tiến hành việc loại bỏ vùng nền còn lại trong khung ảnh nhằm giảm thiểu ảnh hưởng của các vùng này lên mô hình đặc trưng của đối tượng. Tiếp theo, khung ảnh của đối tượng được chia nhỏ thành các mảnh ghép cục bộ và được đặc tả bởi các thông tin về màu sắc, gradient. Bài toán tái nhận dạng người khi này được giải quyết thông qua quá trình tìm kiếm và so sánh các mảnh ghép cục bộ, nhằm ước lượng độ tương đồng giữa các đối tượng cần truy vấn.

Bài báo được tổ chức như sau: mục 2 trình bày các nghiên cứu liên quan đến việc tái nhận dạng đối tượng, phân tích các ưu và nhược điểm của các phương pháp đã đề xuất. Từ các phân tích ở Mục 2, chúng tôi đề xuất trong Mục 3 chuỗi xử lý bài toán tái nhận dạng đối tượng. Mục 4 trình bày các kết quả mô phỏng. Mục 5 là phần kết luận và hướng phát triển.

2 CÁC NGHIÊN CỨU LIÊN QUAN

Có rất nhiều nghiên cứu liên quan đến việc tái nhận dạng người trong hệ thống nhiều camera giám sát, trong đó nổi bật là công bố (Farenzena *et al.*, 2010) phát triển nghiên cứu của nhóm Bazzani *et al.* với đề xuất **Symmetry-Driven Accumulation of Local Features (SDALF)**. Giải thuật đề xuất kết hợp nhiều nguồn thông tin: thông tin về màu sắc toàn cục sử dụng histogram trong không gian HSV, thông tin về phân bố không gian được ước lượng bằng Maximally Stable Color Regions (MSCR) và thông tin về những vùng ảnh có mật độ thông tin tập trung. Tác giả đề xuất trước hết tách người ra khỏi phần nền trong khung ảnh, sau đó tiến hành phân chia các miền đối xứng và bất đối xứng nhằm

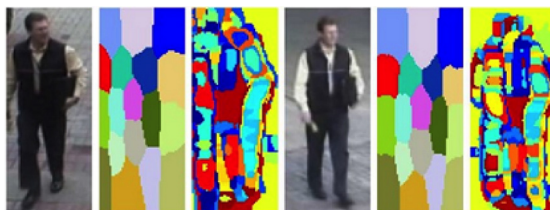
tăng hiệu quả trong việc xây dựng mô hình đặc trưng (Hình 1).



Hình 1: Minh họa mô hình SDALF

Farenzena et al., 2010

X. Wang *et al.*, 2007 đề xuất sử dụng histogram của các gradient có hướng, kết hợp với thông tin về màu sắc để mô tả các đặc trưng của đối tượng cần theo vết. Phương pháp này tỏ ra hiệu quả đối với tập dữ liệu của các đối tượng có sự thay đổi nhỏ về góc nhìn. Giải thuật dựa trên việc phân đoạn khung ảnh của đối tượng thành nhiều vùng bằng cách sử dụng các tiêu chí về hình thái (bộ từ điển hình thái đã được học trước) hoặc tiêu chí về diện mạo bên ngoài. Các mô tả này hợp thành một ma trận thể hiện mối quan hệ không gian về màu của những vùng này như trong Hình 2.

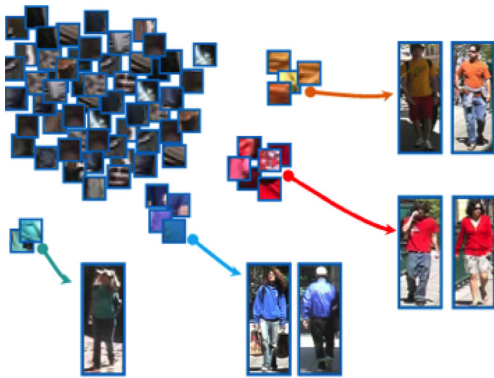


Hình 2: Minh họa cơ sở dữ liệu và kết quả phân đoạn giữa hai hình (từ trái qua phải: hình gốc, kết quả phân đoạn dựa trên hình thái, kết quả phân đoạn dựa trên diện mạo)

X. Wang et al. 2007

R. Zhao *et al.*, 2013 đề xuất phân chia khung ảnh đối tượng thành các mảnh ghép cục bộ (local patch) (Hình 3), từ đó xây dựng ma trận trọng số tương ứng với các vùng nổi bật của đối tượng, nhằm tăng cường độ tin cậy cho quá trình tái nhận dạng. Phương pháp này có ưu điểm là tận dụng được các đặc trưng cục bộ của các đối tượng có bề

ngoài khá giống nhau để phục vụ cho việc tái nhận dạng.

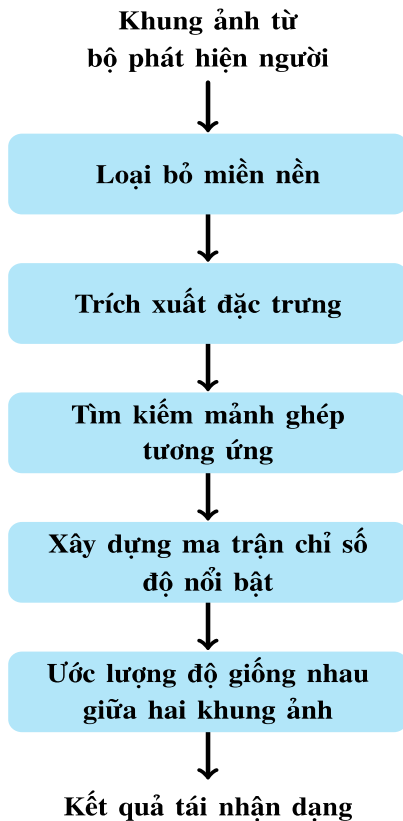


Hình 3: Minh họa miếng ghép cục bộ (local patch) thu được từ các đối tượng

R. Zhao et al., 2013

3 GIẢI THUẬT ĐỀ XUẤT

3.1 Tổng quan



Hình 4: Chuỗi xử lý bài toán tái nhận dạng người

Trong mục này, chúng tôi trình bày chuỗi xử lý đề xuất để giải quyết bài toán tái nhận dạng người. Trước hết, phần nền trong khung ảnh các đối tượng

được loại bỏ nhờ phương pháp đường biên động. Khung ảnh mới sau khi đã loại bỏ miền nền được chia thành các mảnh ghép cục bộ. Đặc trưng về hình dáng bên ngoài của đối tượng truy vấn được đặc tả bởi các vector đặc trưng tính được từ các mảnh ghép này. Độ tương đồng giữa hai khung ảnh truy vấn được ước lượng thông qua quá trình tìm kiếm mảnh ghép tương ứng và ma trận chỉ số đánh giá độ nổi bật của các đặc điểm hình dáng bên ngoài của đối tượng. Chuỗi xử lý có thể được tóm tắt trong lưu đồ ở Hình 4.

3.2 Tách người ra khỏi nền

Trong phần này, chúng tôi ứng dụng giải thuật phân đoạn sử dụng đường bao động được đề xuất trong (Shawn Lankton et al., 2008) để tách đối tượng cần truy vấn ra khỏi nền. Các bước xử lý chính cho phần tách nền có thể được tóm tắt như trong Hình 5.

Gọi C là đường biên khép kín được biểu diễn bởi tập hợp mức 0 của hàm khoảng cách có dấu ϕ :

$$C = \{x | \phi(x) = 0\} \quad (1)$$

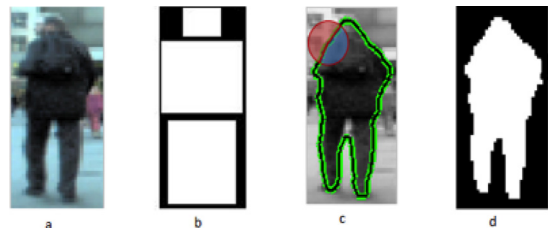
Miền trong của đường biên khép kín C được ước lượng xấp xỉ với hàm Heaviside đã được làm trơn:

$$H[\phi(x)] = \begin{cases} 1, & \phi(x) < -\varepsilon \\ 0, & \phi(x) > \varepsilon \\ S[\phi(x)], & |\phi(x)| < \varepsilon \end{cases} \quad (2)$$

$$\text{với } S[\phi(x)] = \frac{1}{2} \left(1 + \frac{\phi(x)}{\varepsilon} + \frac{1}{\pi} \sin\left(\frac{\pi\phi(x)}{\varepsilon}\right) \right).$$

Tương tự, miền ngoài của C được định nghĩa là $(1 - H[\phi(x)])$.

Miền lân cận quanh đường biên được định nghĩa bởi đạo hàm của hàm $H[\phi(x)]$:



Hình 5: Từ trái qua, (a) Ảnh gốc, (b) Mặt nạ ảnh được định nghĩa trước, (c) Mô tả quá trình tính toán đường biên động, (d) Mặt nạ sau khi trích xuất

$$\delta[\phi(x)] = \begin{cases} 1, & \phi(x) = 0 \\ 0, & |\phi(x)| < \varepsilon \\ D[\phi(x)] & |\phi(x)| \geq \varepsilon \end{cases} \quad (3)$$

$$\text{với } D[\phi(x)] = \frac{1}{2\pi} \left(1 + \cos\left(\frac{\pi\phi(x)}{\varepsilon}\right) \right).$$

Gọi y là biến không gian thứ hai, độc lập với biến x . Cả hai biến này đều dùng để biểu diễn một điểm nào đó trong miền không gian của ảnh. Hàm đặc trưng $B(x, y)$ dùng để đánh dấu các vùng cục bộ có dạng hình tròn với tâm là x và bán kính bằng r được định nghĩa như sau:

$$B(x, y) = \begin{cases} 1, & |x - y| < r \\ 0, & \text{Khác} \end{cases} \quad (4)$$

Hàm năng lượng E trong vùng cục bộ định nghĩa bởi $B(x, y)$ được định nghĩa như sau:

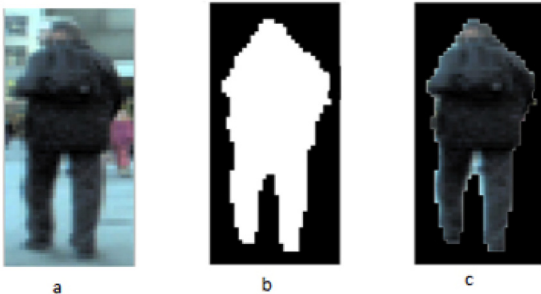
$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x, y) F(I(y), \phi(y)) dy dx \quad (5)$$

với F là hàm năng lượng mô tả mức độ giống nhau và khác nhau giữa các pixel trong các vùng cục bộ dọc đường biên. Các dạng khác nhau của hàm F được định nghĩa chi tiết như trong nghiên cứu (Shawn Lankton et al., 2008).

Để áp dụng giải thuật phân đoạn bằng đường biên động vào việc tách miền nền trong khung ảnh người, một mặt nạ ảnh thô tương ứng với ba vùng đầu, thân và chân của người được khởi tạo, làm cơ sở cho các vòng lặp tìm kiếm. Giải thuật phân đoạn bằng đường biên động sau đó được áp dụng để có được mặt nạ M ứng với vùng đối tượng. Ảnh loại bỏ nền I_M thu được thông qua phép nhân chập ma trận:

$$I_M = I \otimes M \quad (6)$$

Hình 6 trình bày kết quả của quá trình tách đối tượng ra khỏi nền.



Hình 6: Tách đối tượng ra khỏi ảnh nền

3.3 Trích xuất các đặc trưng cục bộ

Để mô tả hình dáng bên ngoài của các đối tượng cần truy vấn, khung ảnh của đối tượng sau khi đã loại bỏ phần nền trước hết sẽ được chia thành các mảnh ghép cục bộ (local patch), với mỗi mảnh ghép có kích thước là $k \times l$ pixel. Mỗi mảnh ghép tiếp theo sẽ được trích hai đặc trưng là histogram trong không gian màu LAB và đặc trưng SIFT. Vector tính được từ hai loại đặc trưng này sẽ được ghép lại thành một vector tổng quát, đặc tả thông tin của từng mảnh ghép, và được gọi là vector đặc trưng dCOLORSIFT (R. Zhao et al., 2013).

3.4 Tìm kiếm mảnh ghép tương ứng

Do góc nhìn của các camera giám sát là khác nhau và người trong quá trình di chuyển sẽ có sự thay đổi về hình dáng bên ngoài, vì vậy các miền tương ứng giữa hai khung ảnh của cùng một đối tượng sẽ không chính xác ở cùng một vị trí. Bước xử lý này sẽ thực hiện việc tìm kiếm các mảnh ghép “tương ứng” trong một miền linh động xung quanh vị trí của mảnh ghép truy vấn. Mảnh ghép tương ứng được xem là mảnh ghép có độ giống nhất với mảnh ghép đem đi truy vấn.

Gọi $\mathbf{x}_{m,n}^{A,p}$ là vector đặc trưng trích từ mảnh ghép ở hàng m và cột n trong ảnh thứ p thu được từ camera A . Gọi $\mathcal{T}^{A,p}(m)$ là tập hợp các mảnh ghép ở hàng thứ m :

$$\mathcal{T}^{A,p}(m) = \{ \mathbf{x}_{m,n}^{A,p} | n = 1, 2, \dots, N \} \quad (7)$$

Tất cả các mảnh ghép trong tập hợp $\mathcal{T}^{A,p}(m)$ có cùng một tập tìm kiếm mảnh ghép tương ứng trong ảnh thứ q thu được từ camera B :

$$\mathcal{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q}) = \mathcal{T}^{B,q}(m), \quad \forall \mathbf{x}_{m,n}^{A,p} \in \mathcal{T}^{A,p}(m) \quad (8)$$

với $\mathbf{x}^{B,q}$ là tập hợp các mảnh ghép trong ảnh q thu được từ camera B . Trong trường hợp lý tưởng, tập S là hàng thứ m của ảnh q ở camera B . Tuy nhiên, việc tìm kiếm theo hàng tương ứng thường không mang lại hiệu quả cao vì sự thay đổi góc nhìn của camera giám sát, cũng như sự thay đổi về hình dáng của người khi di chuyển. Để đối phó với các biến thể không gian, việc tìm kiếm được mở rộng ra trên các hàng liền kề với hàng cần tìm kiếm, việc tìm kiếm này được mô tả theo công thức sau:

$$\hat{\mathcal{S}}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q}) = \{ \mathcal{T}^{B,q}(b) | b \in N(m) \}, \quad (9)$$

$$\forall \mathbf{x}_{m,n}^{A,p} \in \mathcal{T}^{A,p}(m)$$

với $N(m) = \{m-l, \dots, m, \dots, m+l\}$, $m-l \geq 0$ và $m+l \leq M$, M là số hàng của khung ảnh p, q . Giá trị l xác định số hàng linh động cho việc tìm kiếm.

Đối với mỗi mảnh ghép có vector đặc trưng $\mathbf{x}_{m,n}^{A,p}$ trong khung ảnh p của camera A , chúng tôi thực hiện tìm kiếm mảnh ghép tương ứng trong khung ảnh q của camera B , nghĩa là tìm kiếm vector gần nhất trong tập hợp $\hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$. Hàm ước lượng độ giống nhau của hai vector đặc trưng được định nghĩa thông qua hàm Gaussian:

$$s(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{(\|\mathbf{x} - \mathbf{y}\|_2)^2}{2\sigma^2}\right) \quad (10)$$

với σ là tham số cho trước.

3.5 Ước lượng chỉ số đánh giá độ nổi bật

Ở phần này chúng tôi trình bày hai phương pháp được áp dụng là k -Nearest Neighbors (k -NN) (S. Byers và A. Raftery, 1998) và One Class Support Vector Machines (OCSVM) (K. Heller *et al.*, 2003) để xây dựng ma trận chỉ số đánh giá độ nổi bật của các đặc trưng hình dáng bên ngoài đối tượng.

3.5.1 k -NN

Gọi N_r là số lượng các ảnh trong tập hợp ảnh tham chiếu cho quá trình tái nhận dạng. Sau khi xây dựng vector đặc trưng (dCOLORSIFT) và thực hiện việc tìm kiếm các mảnh ghép tương ứng giữa ảnh truy vấn và các ảnh trong tập tham chiếu, ứng với mỗi mảnh ghép truy vấn $\mathbf{x}_{m,n}^{A,p}$, ta có được N_r mảnh ghép tương ứng tìm được từ N_r ảnh tham chiếu. Tập hợp các mảnh ghép này, tạm gọi là $X_m(\mathbf{x}_{m,n}^{A,p})$, được định nghĩa theo công thức sau:

$$X_m(\mathbf{x}_{m,n}^{A,p}) = \left\{ X \mid \arg \max_{\hat{x} \in \hat{S}_{p,q}} S(\mathbf{x}_{m,n}^{A,p}, \hat{x}), q = 1, \dots, N_r \right\} \quad (11)$$

với $\hat{S}_{p,q} = \hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$ là tập tìm kiếm lân cận trong phương trình (9) và s là hàm ước lượng độ giống nhau định nghĩa trong phương trình (10).

Khoảng cách KNN (S. Byers và A. Raftery, 1998) được sử dụng để ước lượng chỉ số đánh giá độ nổi bật (salience score):

$$\text{score}_{knn}(\mathbf{x}_{m,n}^{A,p}) = D_k(X_{nn}(\mathbf{x}_{m,n}^{A,p})) \quad (12)$$

với D_k là khoảng cách giữa mảnh ghép truy vấn và mảnh ghép tương ứng gần nhất thứ k .

3.5.2 OCSVM

Giải thuật thứ hai được áp dụng trong phần này là One Class Support Vector Machines (OCSVM) (K. Heller *et al.*, 2003). Ý tưởng cơ bản của OCSVM là tìm kiếm một mặt siêu cầu biểu diễn được hầu hết các điểm dữ liệu trong không gian đặc trưng và có kích thước nhỏ nhất có thể. Mặt siêu cầu này tìm được thông qua việc giải bài toán tối ưu với hàm quyết định $f(\mathbf{X})$ được định nghĩa trong (R. Zhao *et al.*, 2013).

Chỉ số đánh giá độ nổi bật trong trường hợp này được ước lượng như sau:

$$\text{score}_{ocsvm}(\mathbf{x}_{m,n}^{A,p}) = d(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^*) \quad (13)$$

với $\mathbf{x}^* = \arg \max_{\mathbf{x} \in X_m(\mathbf{x}_{m,n}^{A,p})} f(\mathbf{x})$, d là khoảng cách

Euclidean giữa các vector đặc trưng.

3.6 Tái nhận dạng đối tượng

Phần này trình bày bước cuối cùng trong chuỗi xử lý bài toán tái nhận dạng đối tượng. Xét một cặp ảnh gồm ảnh của đối tượng cần truy vấn (ảnh p) và ảnh trong tập tìm kiếm (ảnh q). Ứng với mỗi mảnh ghép $\mathbf{x}_{m,n}^{A,p}$, ta tìm được mảnh ghép tương ứng trong miền tìm kiếm $\hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$ là:

$$\mathbf{x}_{i,j}^{B,q} = \arg \max_{\hat{x} \in \hat{S}_{p,q}} S(\mathbf{x}_{m,n}^{A,p}, \hat{x}) \quad (14)$$

Ảnh của đối tượng trong tập ảnh tìm kiếm có độ tương đồng cao nhất với ảnh truy vấn được xác định như sau:

$$q^* = \arg \max_q \text{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q}) \quad (15)$$

với hàm Sim là hàm đo độ giống nhau giữa hai đối tượng trong khung ảnh p và q . Hàm Sim được ước lượng như sau:

$$\text{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q}) = \sum_{m,n} \frac{\text{score}(\mathbf{x}_{m,n}^{A,p}) \cdot s(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}_{i,j}^{B,q}) \cdot \text{score}(\mathbf{x}_{i,j}^{B,q})}{\alpha + |\text{score}(\mathbf{x}_{m,n}^{A,p}) - \text{score}(\mathbf{x}_{i,j}^{B,q})|} \quad (16)$$

với α là tham số cho trước, $score$ có thể là giá trị $score_{knn}$ hoặc $score_{ocsvm}$.

4 KẾT QUẢ VÀ THẢO LUẬN

4.1 Cơ sở dữ liệu

Chúng tôi sử dụng cơ sở dữ liệu VIPeR (Viewpoint Invariant Pedestrian Recognition) (D. Gray *et al.*, 2007) để làm cơ sở dữ liệu mô phỏng. Đây là cơ sở dữ liệu được sử dụng phổ biến, rộng rãi để kiểm tra sự hiệu quả của các phương pháp tái nhận dạng đối tượng. Cơ sở dữ liệu này bao gồm 632 ảnh của các đối tượng thu được từ hai camera có góc quay và độ chiếu sáng khác nhau. Các ảnh này đều có kích thước là 124×48 . Các hình ảnh thu được từ camera A có góc nhìn từ 0^0 đến 90^0 , còn camera B có góc nhìn từ 90^0 đến 180^0 .

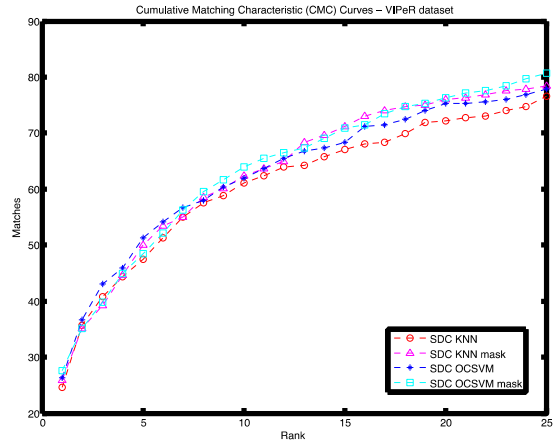
4.2 Cách thức thực hiện mô phỏng

Theo như cách thức được đề xuất trong (Gray, Douglas và Hai Tao, 2008), chúng tôi lấy tùy ý 316 cặp ảnh dùng cho mục đích huấn luyện và 316 cặp còn lại dùng cho kiểm tra. Trong thí nghiệm này, các ảnh ở camera A được sử dụng như là một tập hợp các đối tượng cần được nhận dạng, còn các ảnh ở camera B được sử dụng như là một tập các đối tượng đã được định danh. Độ chính xác của bài toán nhận dạng phụ thuộc vào cấu trúc của tập hợp A và B . Nếu A và B càng chứa nhiều thông tin đặc trưng riêng cho từng đối tượng để so sánh thì độ chính xác càng cao, và dĩ nhiên đi kèm với đó, độ phức tạp cũng sẽ tăng lên. Kết quả tái nhận dạng được biểu diễn bằng đường cong Cumulative Matching Characteristic (CMC) (X. Wang *et al.*, 2007).

4.3 Kết quả mô phỏng

Hình 7 biểu diễn kết quả các đường CMC của các giải thuật SDC_KNN, SDC_OCSVM (R. Zhao *et al.*, 2013), và hai giải thuật do nhóm nghiên cứu đề xuất là SDC_KNN_mask, và SDC_OCSVM_mask (T. Toàn, 2015). Nhìn vào đồ thị này ta thấy rằng giải thuật đề nghị có cái tiến tương đối so với giải thuật gốc ban đầu.

Bảng 1 so sánh kết quả nhận dạng với các phương pháp khác nhau: LMNN (W. Zheng, 2011), PCCA (A. Mignon và F. Jurie, 2012), SDC_kNN (R. Zhao *et al.*, 2013), SDC_OCSVM (R. Zhao *et al.*, 2013), SDALF (Farenzena *et al.*, 2010), và phương pháp cải tiến của (R. Zhao *et al.*, 2013), có kết hợp mặt nạ ảnh là SDC_kNN_mask và SDC_OCSVM_mask (T. Toàn, 2015).



Hình 7: Kết quả tái nhận dạng người thu được với bộ cơ sở dữ liệu VIPeR

Bảng 1: So sánh kết quả tái nhận dạng người với các phương pháp khác nhau trên bộ cơ sở dữ liệu VIPeR

Phương pháp	$R = 1$	$R = 5$	$R = 10$	$R = 20$
LMNN	6.23	19.65	32.63	52.25
PCCA	19.27	48.89	64.91	80.28
SDALF	19.87	38.89	49.37	65.73
SDC_kNN	24.68	47.47	61.08	72.15
SDC_OCSVM	26.27	51.27	62.03	75.32
SDC_kNN_mask	25.95	50.00	62.34	75.95
SDC_OCSVM_mask	27.53	48.42	63.92	76.27

5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, chúng tôi trình bày chuỗi xử lý bài toán tái nhận dạng đối tượng với hai bước xử lý chính: tách người ra khỏi miền nền còn lại trong khung ảnh và thực hiện việc tìm kiếm ảnh tương đồng để xác định đối tượng cần truy vấn. Kết quả thực nghiệm cho thấy mô hình đề xuất cải thiện được một phần hiệu suất nhận dạng đối tượng so với một số phương pháp khác.

Một số hướng phát triển có thể xem xét để nâng cao hiệu quả của bài toán tái nhận dạng như khảo sát kích thước mảnh ghép cục bộ nhằm chọn ra thông số kích thước tối ưu, hay kiểm nghiệm một số đặc trưng khác trong quá trình trích xuất đặc trưng trên các mảnh ghép cục bộ.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Quỹ phát triển khoa học và công nghệ quốc gia (NAFOSTED) trong đề tài mã số 102.99-2013.36.

TÀI LIỆU THAM KHẢO

1. B. Prosser, W. Zheng, S. Gong, T. Xiang, and Q. Mary, 2010. Person re-identification by support vector ranking, in *Proceedings of the British Machine Vision Conference (BMVA)*.
2. Hirzer Martin, Csaba Beleznai, Peter M. Roth, Horst Bischof, 2011. Person re-identification by descriptive and discriminative classification, in *Proceedings of 17th Scandinavian Conference on Image Analysis (SCIA)*, pp. 91 - 102.
3. Mert Dikmen, Emre Akbas, Thomas S. Huang, and Narendra Ahuja, 2011. Pedestrian recognition with a learned metric, in *Proceedings of the 10th Asian Conference on Computer Vision (ACCV)*, Volume Part IV, pp. 501 - 512.
4. W. Zheng, S. Gong, and T. Xiang, 2011. Person re-identification by probabilistic relative distance comparison, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 649 - 656.
5. Martin Hirzer, Peter M. Roth, Martin Köstinger, Horst Bischof, 2012. Relaxed pairwise learned and metric for person re-identification, in *Proceedings of the 12th European Conference on Computer Vision (ECCV)*, Vol. Part VI, pp. 780 - 793.
- A. Mignon and F. Jurie, 2012; PCCA: A new approach for distance learning from sparse pairwise constraints, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2666-2672.
6. M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, 2010. Person re-identification by symmetry-driven accumulation of local features, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2360-2367.
7. X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, 2007. Shape and appearance context modeling, in *Proceedings of International Conference on Computer Vision (ICCV)*, 2007, pp. 1-8.
8. R. Zhao, W. Ouyang, and X. Wang, 2013. Unsupervised salience learning for person re-identification, in *proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3586-3593.
9. Shawn Lankton, Allen Tannenbaum, 2008. Localizing Region-Based Active Contours, In *IEEE Transactions on Image Processing*, Vol. 17, no. 11, november 2008, pp. 2029 - 2039.
10. S. Byers and A. Raftery, 1998. Nearest-neighbor clutter removal for estimating features in spatial point processes, *Journal of the American Statistical Association*, No. 442, 1998, pp. 577-584.
11. K. Heller, K. Svore, A. Keromytis, and S. Stolfo, 2003. One class support vector machines for detecting anomalous windows registry accesses, In *Workshop on DataMining for Computer Security (DMSEC)*, pp. 2-9.
12. Y. Chen, X. Zhou, T. Huang, 2001. One-class SVM for learning in image retrieval, in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2001, pp. 34-37.
13. D. Gray, S. Brennan, and H. Tao, 2007, Evaluating appearance models for recognition, reacquisition, and tracking, in *Proceedings of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, vol. 3, no. 5.
14. Gray, Douglas, and Hai Tao, 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features, in *Proceedings of 10th European Conference on Computer Vision (ECCV)*, pp. 262 - 275.
15. Trần Thanh Toàn, 2015. Tái nhận dạng người trong các hệ thống nhiều camera giám sát, *Luận văn thạc sĩ ngành Kỹ thuật điện tử, Trường Đại học Bách Khoa, Đại học Quốc gia Tp Hồ Chí Minh.*